Drawing Samples for the Longitudinal Study of Entrepreneurial Groups from Process-Generated Data: A Proposal Based on the German Register of Companies

Weinhardt, Michael; Stamm, Isabell

Drawing Samples for the Longitudinal Study of Entrepreneurial Groups from Process-Generated Data: A Proposal Based on the German Register of Companies

Michael Weinhardt & Isabell Stamm *

Abstract: »*Stichprobenziehung für die längsschnittliche Betrachtung unternehmerischer Gruppen aus prozess-produzierten Daten: Ein Vorschlag auf Basis des deutschen Handelsregisters*«. The growing interest in entrepreneurial groups as collective actors of entrepreneurship raises questions of how and with what kind of data this unit of analysis can be studied. While sampling and access to data on individual entrepreneurs (self-employed) or their business ventures (formal firms) rest upon established routines, a methodological discussion about identifying and sampling entrepreneurial groups is still in its infancy. In this article, we look at process-generated data as a potential linchpin to study entrepreneurial groups. More particularly, this article critically reflects upon the opportunities and challenges of the German Commercial Registry (CR) to function as a sampling frame and data source for an examination of entrepreneurial groups. This reflection includes a discussion about the key characteristics of entrepreneurial groups in order to derive minimal criteria that the data needs to provide, an evaluation of the CR following a data source study approach, and finally an assessment of the error proneness of this data and its consequences for the study of entrepreneurial groups. On this basis, we propose a sampling strategy of entrepreneurial groups with CR data. As such, this article contributes to a general methodological discussion of process-generated data, as it extends and practically applies the concept of a data source study. It also contributes to a methodological discussion about entrepreneurial groups as it offers a procedure to deal with varying group boundaries and the intertwinement of group and business activity typical for this social unit of analysis.

Keywords: Sampling frame, data source study, commercial registry, data quality, Handelsregister.

* Isabell Stamm, Technical University Berlin, Department of Sociology; Fraunhofer Str. 33-36, 10587 Berlin, Germany; isabell.stamm@tu-berlin.de.
Michael Weinhardt, Technical University Berlin, Department of Sociology; Fraunhofer Str. 33-36, 10587 Berlin, Germany; michael.weinardt@tu-berlin.de.

Historical Social Research 44 (2019) 4, 186-221 | published by GESIS
DOI: 10.12759/hsr.44.2019.4.186-221

1. Introduction

An increasing number of scholars agree that entrepreneurship requires not a single actor, but involves multiple actors collectively (Harper 2008; Schjoedt et al. 2013). Through their joint investment of time, effort, and money, multiple individuals contribute to the entrepreneurial process (Gartner 1988; Moroz and Hindle 2012). A practice perspective on entrepreneurship pertinently coined the term "entrepreneuring" to underline the continuous and bundled nature of this special economic activity (Johannisson 2014; Steyaert 2007), which does not exhaust itself in the foundation of an organization (e.g., a business or an association), but instead involves constant development of entrepreneurial projects (Shane 2000). For the time of their joint engagement in discovering, evaluating, and exploiting opportunity, they form an entrepreneurial group (Klotz et al. 2014; Ruef 2010). These groups can attach*, detach, and* re-attach themselves to entrepreneurial projects.

Such a relational perspective identifies the entrepreneurial group as a distinct, additional unit of analysis: Entrepreneurial groups are a small number of individuals bound together by their joint engagement in entrepreneuring activities that form cliques in the social networks of these entrepreneurial individuals (the group members); occupy formal and informal roles in the business organization(s) they are involved in; and are key actors in shaping their respective fields and industries. Acknowledging entrepreneurial groups as relevant actors in entrepreneurship poses new questions about their emergence, development, and dissolution, thereby increasing our understanding of entrepreneuring as a socially embedded process (see Stamm, Discua Cruz, and Cailleut 2019).

To advance an understanding of how entrepreneurial group processes unfold over time, as opposed to a static approach focusing on how variables co-vary at a single point in time, the availability of retrospective, longitudinal, process-generated data is crucial. In addition, studying the development of entrepreneurial groups requires a heightened complexity of the research design. A sampling strategy is needed which allows drawing samples from past years. In this way it becomes possible to investigate the trajectory of an entrepreneurial group in relation to the involved individuals' formal engagement in an entrepreneurial project and organizational evolution.

However, despite a growing number of empirical studies about various collective actors of entrepreneurship (e.g., on new venture teams (Preller, Breugst, and Patzelt 2015), entrepreneurial families (Leiß 2014), or community entrepreneurship (Nordstrom and Jennings 2015)), both theorizing about entrepreneurial groups and reflecting about the methodological challenges in empirically studying them remains scarce. In addition, those studies that do exist are mostly based on samples from specific industries or types of entrepreneuring activities (such as start-ups), which do not allow for broad conclusions on the

overall phenomenon of entrepreneurial groups (Breugst and Shepherd 2017; Bonardo, Paleari, and Vismara 2010; Block, Hornuf, and Moritz 2018; Koeller and Lechler 2006; Lechler 2001; Klingbeil and Semrau 2016; Brinckmann and Hoegl 2011; Semrau and Sigmund 2012; Talaulicar, Grundei, and Werder 2005). Overall, there is a lack of studies based on representative samples of entrepreneurial groups that might allow us to draw valid conclusions about their prevalence at one point in time as well as their development and changing composition over time (Hellerstedt 2009).

Methodological considerations for the study of entrepreneurial groups are closely linked to conceptual ones, such as the definition of group boundaries, the distinction of group activities from business activities, and the evolution of entrepreneurial groups over time. Clear conceptual definitions are a necessary precondition to drawing valid samples in order to determine and select sampling units correctly and to assess the completeness of all necessary information. Furthermore, the mapping of activities of entrepreneurial groups over time requires a combination of processual and multi-level concepts and analysis, which forms a theoretically and methodologically challenging task (Harper 2008). Both panel surveys as well as ethnographic data allow the mapping of entrepreneurial group dynamics simultaneously. For a retrospective inquiry of long-term developments of entrepreneurial groups, the use of already existing data that was generated in the process of this past development is crucial (Hellerstedt 2009). Such retrospective inquiries must draw samples from the past and gather data on the subsequent trajectories of entrepreneurial groups. Otherwise, if such inquiries rely on entrepreneurial groups only existent today, the sample would be heavily biased towards the "survivors". Hence, the construction of a valid sample requires a data source which allows "going back in time" and selecting entrepreneurial groups to become part of the analysis. Process-generated, official data from commercial and trade registers appear to be a suitable candidate for this endeavor.

In this article, we describe how we conceptualize and operationalize entrepreneurial groups in the course of an ongoing research project entitled "Entrepreneurial Group Dynamics"[1]. As part of this project, we intend to study the long-term development of entrepreneurial groups. For this purpose, we make use of process-generated data on founders and their businesses[2] from different sources, including official register data. As we take a longitudinal perspective, we realize that entrepreneurial groups are distinctive entities that need to be treated separately from the business organizations they are engaged in. There-

[1] More information on the project, its theoretical foundations, and empirical findings can be found here: <http://www.entrepreneurialgroups.org>.

[2] For the course of this article, we use the term "business" to refer to a business organization that may be emerging (venture), has a legal form (company), and may itself be interconnected to other businesses (business group).

fore, our conceptualization of entrepreneurial groups is new in important ways: First, an entrepreneurial group may be engaged in more than one venture at a time, resulting in multiple businesses. Second, not all group members must be engaged in all businesses simultaneously. As entrepreneurial groups and business organizations cannot be equated, this has serious consequences for the empirical study of these phenomena as it makes the identification of our unit of inquiry much more difficult. Also, the retrospective study design necessitates the retroactive construction of samples from previous years, which complicates the sampling process even further.

To address these challenges, we discuss the German Register of Companies ("Handelsregister"; CR in the following) as a potential data source to construct a valid sampling frame for entrepreneurial groups, dating back up to 20 years or more. The CR offers a census-like database of all current businesses in Germany in which newly founded businesses are required by law to register in order to do business. In addition to the fact that it allows for the identification of all new businesses in a given year, it lists all the chief executives of a business and it allows the identification of shareholders of capital stock companies. This completeness of information permits us to check for ties of all individuals engaged in the ownership and management of one newly established business with other individuals and companies and therefore allows us to draw samples of entrepreneurial groups.

Despite the overall suitability of the usage of CR data as a sampling source, we also discuss potential pitfalls of this approach by investigating several coverage and data quality issues of the German CR. We particularly focus on the biases that result from the data production and archiving process in such data and their effect on measurement and sample quality (representation). For this purpose, we follow the framework of a critical "data source study" that has been put forward by Bick and Müller (1984) and advanced by Baur (2004, 2009). The framework allows for a structured assessment of the German CR in terms of potential threats to validity. Based on this in-depth data exploration exercise, and in spite of certain instances of under-coverage and over-coverage existing at both the organization level as well as the group level, we find the German CR to be a suitable source for sampling entrepreneurial groups over time. As all analyses and derived findings will only be as good as the quality of the sample they are based on, this paper is an important contribution to the methodological discussion about research on entrepreneurial groups.

In the following, we start by describing the conceptual challenges in defining entrepreneurial groups and distinguish them from business organizations. We then describe our proposed strategy to sample entrepreneurial groups from process-generated data and derive the required information for this proposal. Next, we introduce the German CR as a source of process-generated data, including its overall set-up and content, and show how it matches the requirements to serve as a data source for the proposed strategy. To discuss its useful-

ness as a sampling frame, we evaluate the quality of register data according to a data-source framework. This forms the basis for a discussion of the quality, usefulness, and trustworthiness of the CR data as a source to construct valid samples for the long-term investigation of entrepreneurial groups.

2. Entrepreneurial Groups: Previous Research and Conceptual Considerations

Despite early recognition of the social embeddedness of entrepreneurship, research on entrepreneurial groups has remained scarce for a long time.[3] An unease with an individualistic interpretation of entrepreneurship can be traced back to Schumpeter, who noted that "the entrepreneurial function need not be embodied in a physical person and in particular in a single physical person" (Schumpeter 1965, 51). Nevertheless, the myth of the entrepreneur as a lone actor has been reflected in the entrepreneurship literature, creating a sizeable gap that still continues to exist in the conceptual and empirical literature on entrepreneurial groups (Cooney 2005). A milestone in the emergence of a relational perspective on entrepreneuring was marked by the research agenda suggested by Kamm et al. (1990). Based on a narrow understanding of entrepreneurship as new venture creation, they define entrepreneurial groups "as two or more individuals who jointly establish a business in which they have an equity (financial) interest" (Kamm et al. 1990, 7) and call for defining the dimensions of such groups, the challenges of assembling groups, and the identification of success factors. Since then, the number of contributions dedicated to studying joint engagement in entrepreneuring has grown tremendously (Klotz et al. 2014). These studies are evidence that a large number of businesses are indeed founded by groups (Ruef, Aldrich, and Carter 2003; Ruef 2010). Among the topics that have received the largest interest among scholars are group composition and member resources (such as prior experience or social capital). These studies provide the basis for a highly controversial debate about the effects of a homogeneous or heterogeneous group composition on the success of a business venture (Ruef 2010; Ruef, Aldrich, and Carter 2003).

Despite this recent upturn in research on entrepreneurial groups, key issues remain unresolved, as studies have so far produced mixed and inconsistent findings. This may well be due to a lack of quality of samples in those studies on entrepreneurial groups. With the exception of Ruef (2010), existing research has paid little attention to the quality of the sample in order to draw inferences

[3] See also the editorial to this special issue for a discussion of the existing literature on entrepreneurial groups, their conceptual definition, and, respectively, the lack thereof in previous research (Stamm, Discua Cruz, and Cailluet 2019).

to the population. This needs to change if we want to produce conclusive evidence on the role that groups play in the entrepreneurial process.

One key obstacle in the empirical study of entrepreneurial groups is the lack of conceptual clarity. We suggest that at least two areas are crucial: the definition of group boundaries (a) as well as a distinction between group and business activity (b). In the following, we will address these issues by summarizing existing research and pointing towards analytical touchstones. This shall serve to guide pragmatic decisions necessary for the empirical study of entrepreneurial groups over time.

2.1 Defining Group Membership and Boundaries

An entrepreneurial group can be viewed as a clique or a coalition within a larger network (Taylor 1999; Ruef 2010). In comparison to the broad network relations, the entrepreneurial group is a small group characterized by stronger and relatively stable bonds. The group members jointly follow an entrepreneurial project under conditions of uncertainty, with their activities creating an interdependence between the members (Harper 2008). Over time, who is in and out of the group may change, which makes the definition of group borders a pressing conceptual issue.

One way of approaching this issue is by defining group membership in terms of commitment to an entrepreneurial project following a tradition of defining group membership by self-identification (Schäfers 1994). An individual is classed as a group member when he or she ascribes to a shared vision of implanting future shaping activities. These visions can circulate around producing a specific product, delivering a service, or creating a new model. They can also focus on the individual or collective generation of income, wealth, or well-being, viewing the materialization of an idea as a means to a higher end (Taylor 1999; Berrone, Cruz, and Gomez-Mejia 2012). In other words, the commitment to a project does not require that all individuals have the same intentions and motives (Schjoedt and Kraus 2009). Small group research suggests that self-identification as a group member, and hence commitment to a project, may differ from the ascription of membership through others (Tajfel 1982); while some members may feel attached to a group, other members or outsiders may not view them as part of the group. Given these discrepancies, the question arises of whether a reversion to some form of official acknowledgement in order to consider someone a member of an entrepreneurial group would be beneficial.

An alternative way of defining group boundaries is by defining group membership in terms of actual engagement in the entrepreneurial project. Ruef (2010) suggests that this engagement may include the investment of time, money, or effort. This suggestion encompasses such heterogeneous forms of engagement, ranging from full-time work to occasional helping out, from taking

private financial risks to professional investments, and from emotional support to advice. Some group members may engage in multiple ways and by their continuous involvement in the project, while the engagement of others is highly specialized and short-term. Some forms of this engagement institutionalize to set roles within the entrepreneurial group, some roles may stay informal, while others are formalized within the respective businesses, drawing from predefined roles in the entrepreneurial field or by legal registration (e.g., an owner with or without voting rights, a chief executive officer, a business angel investor). Defining a group boundary on the level of institutionalized roles decreases differences between self-identification and ascription. It also largely reduces the studied group phenomenon to small groups rather than social groups (e.g., workers, students).

The criteria used to define a group's boundary will yield a countable group size at a given point in time. In sociology and social psychology, both the lower limit and the upper limit of small group size – which distinguishes the small group from social groups, groupings, or a crowd – are heatedly debated (Schäfers 1999; Simmel 1992). In this line of research, the following size categories appear to be the most prevalent: dyads (such as entrepreneurial spouses or copreneurs), triads and small groups (up to 20 members). It is important to note that groups of different sizes will exhibit different levels of changes and dynamics.

2.2 Distinguishing Group from Business Activity

The attempt to study a group's trajectory in relation to an organization rests upon the assumption that both are distinct entities. This assumption, however, needs to be reflected upon with some reservations. After all, the emergence of the organization largely rests upon the activities that group members engage in; the business created is thus a patterned outcome of crystallized group relationships (Taylor 1999). In the further course of the development, a close intertwinement between organization and group activities continues to exist. While the group may imprint rules of action, values and routines in the organization (Marquis and Tilcsik 2013), the organizational structure may limit or enable group activity. The intertwinement between group and organization may vary strongly, ranging from coincidental to loosely coupled, depending on the ratio between group and organizational size.

In group research and organization studies, a debate about the differentiation between both units of analysis has a long standing tradition (Neidhardt 1983). For example, Claessens (1983) described how groups can experience internal and external pressures to organize. Homans (1972) suggested that in groups, actions are personalized, in distinction to institutionalized and depersonalized roles and rules for action in organizations. In this discussion, it is implied that a group is either a partial unit within one organization or the group is separate

from an organization. With entrepreneurial groups, the variety is larger. The group may be entangled with multiple organizations, thereby building pyramidal structures, business portfolios, or sequences of entrepreneurial action. This also implies that a group has the capacity to disentangle from an organization, with both entities continuing to exist without each other.

An important piece in this puzzle appears to be the adapted legal entity of a firm. The legal form, it can be argued, is no more than a convention of regulation that a set of actors has decided to adopt at a given point in time and, in the following, frames both business and group activity. In Germany, for example, a number of legal forms exist which entrepreneurial groups can choose from when founding, reflecting their arrangement of engagement in the operation of the business, their respective ownership, and agreed-upon voting structure as well as their tolerance for personal liability (e.g., Katz 1993). As the group forms, reforms, and dissolves and/or the business stagnates, declines, or grows, the legal form may be changed, and so too the legal forms available to choose from (e.g., in Germany the Unternehmergesellschaft UG was introduced in 2008 and the gGmbH was introduced in 2013). While the legal form cannot be equated with either the organization (e.g., Taylor 1999) or the entrepreneurial group, the legal form needs to be understood as an institutionalized framework for both the business operation and the relationship between involved group members. In other words, the form, or changes in the form, of a firm may not mirror all underlying activities, but at least allow the drawing of inferences about these.

Following this conceptual discussion of entrepreneurial groups and considering the possibilities and constraints of the process-generated data, we suggest to operationalize group membership by occupying a formalized role – as owners, investors, or executive managers – within at least one of the respective businesses. Hence, we operationalize group membership in terms of group boundaries via functional (and not self-identity) criteria.[4] Being aware that this definition is rather restrictive and does not mirror additional forms of engagement such as advice, labor, or emotional support, we set the lower limit of a group size to two members. We also acknowledge that entrepreneurial groups and the founded companies are separate but linked entities by considering that a group member may hold multiple formal roles across various businesses. Finally, an activity is only considered a group activity when multiple (if not all)

[4] In order to make self-identification a defining criterion for group membership, we would need survey data from each potential group member on this issue. This has two problems: first, we would still need another criterion telling us which people fall in the circle of potential group members who should be asked about their potential self-identification. Second, due to the longitudinal nature of our envisioned study design, retrospective survey questions about self-defined group membership from five or ten years ago is likely to be invalid.

group members are jointly engaged in the respective business; however, not all members need to be engaged in all of the group's business activities at the same time.

3. Sampling Entrepreneurial Groups: Strategy and Data

The conceptual discussion of group membership and the groups' relation to the business organizations is a much-needed clarification on our actual unit of analysis and starting point for the construction of valid samples. We suggest operationalizing entrepreneurial groups as a set of individuals that is continuously engaged in the founding and operating of one or multiple businesses. The engagement of individuals in the coordinated entrepreneurial activities can take the shape of investment, ownership, or executive management. The minimum size of an entrepreneurial group should be set to two. Previous research on entrepreneurial groups has not yet put forward an established and readily available sampling strategy that can be used to draw samples on which to base valid inferences to the population, especially not as defined here. Both the distinction between entrepreneurial groups and businesses – and the fact that one entrepreneurial group may be involved in more than one business organization at a time – pose a big challenge for the sampling of entrepreneurial groups. Identifying newly founded business companies alone will not suffice, but will need to be followed by an effort to establish whether the founders of this particular business are also involved in other companies – and whether those additional companies point to more individuals who may also be group members (while not being engaged in the company originally selected). Based on an extensive search of the literature, we feel safe to conclude that these two aspects have been mostly neglected in previous studies of entrepreneurial groups. In addition, we are ambitious in that we aim for the big picture, targeting all groups of all sizes in the country rather than concentrating on certain regions or specific industries. In the following, we will sketch out a sampling strategy that allows the identification of entrepreneurial groups, taking into account these conceptual complexities and the broad scope of the envisioned sample. From the discussion of this sampling strategy, we then derive the requirements that a data source must meet before we can use it as a sampling frame for entrepreneurial groups.

3.1 A Sampling Strategy for Entrepreneurial Groups

The population we are interested in is the totality of all entrepreneurial groups in Germany according to the definition given above. To draw probability samples, we usually need a *sampling frame* that lists all elements in the target population from which a sample may be drawn (cf. Schnell et al. 2005, 265). It

is most commonly required that, when drawing a sample from the frame, each element must have a known, non-zero probability of being selected (Groves et al. 2011). Usually, the sampling frame must allow for computing the probability of being selected for each element it contains. However, it is important to note that entrepreneurial groups are an entity of interest for researchers and scientists, not immediately for government authorities or other bodies who usually put together lists from which samples can be drawn. As German commercial law allows for a range of collaborative engagements in founding and operating companies – it is not the entrepreneurial group, but rather their individual members or the attached businesses that are subject to taxation or policy regulations. Hence, there is neither a public listing nor official statistical reporting that may be used for generating a sample of entrepreneurial groups (Hellerstedt 2009; Ruef 2010).

To address the challenge of a non-existing list of entrepreneurial groups, we suggest an approach similar to what is called hypernetwork sampling in the literature. This approach has been used in similar situations where there is no sampling frame of the target elements but for other, related elements. Examples include the sampling of employers via employees in so-called linked-employer-employee studies (Kalleberg 1996) or the sampling of protest events via the sampling of protesters first. Ruef (Ruef 2010) also used this approach to sample entrepreneurial groups by first surveying individual entrepreneurs in the Panel Study of Entrepreneurial Dynamics (PSED). While this approach will help us in identifying groups from their individual members, we still need a list of entrepreneurs to begin with.

Step 1: As a starting point of the sampling strategy, we suggest generating a probability sample of newly founded businesses and extracting all individuals involved in entrepreneurial activity. This is a challenge in itself. Overall, the options for sampling companies in general are limited because one simple-to-use, all-encompassing sampling frame is lacking (Hartmann 2017). Past studies are restricted in their scope as they often focus solely on start-ups and lists generated from incubators and accelerators; hence, they only yield a very specialized picture of founding teams. The Federal Employment Agency keeps a list of all organizations that have employees paying social security contributions. However, information on shareholders and executive managers is not directly included and its use is highly restricted due to the confidentiality of the data. Other possibilities, such as lists of members of chambers of commerce and industries (IHKs) or similar entities, are restricted to certain industries and local areas and do not cover the broad picture of entrepreneurial activity in Germany. Another option is using the so-called employee-first method (Kmec 2003), but this would require a large sample of employees first and, hence, is not readily available either. The fact that we need, in particular, newly established companies complicates the task even further. There are local trade registrations (*Gewerbeanzeigen*) which indicate early business activity,

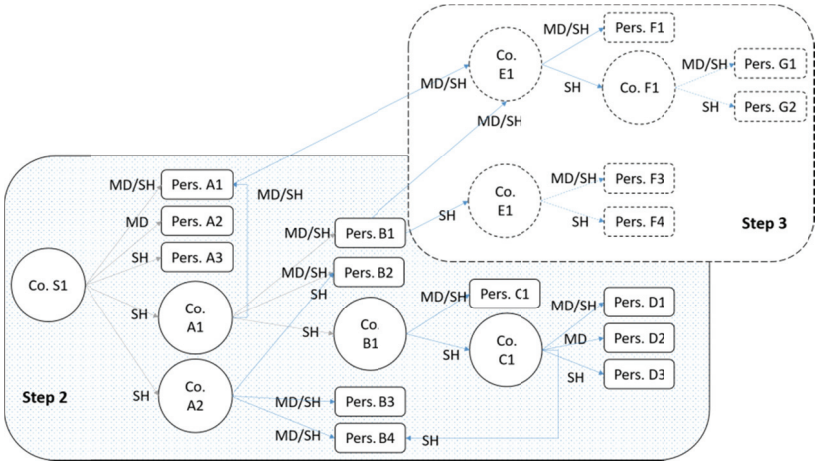
but access is decentralized and highly restricted (Brüderl, Preisendörfer, and Ziegler 1992). Also, they do not include information on the owners and/or shareholders of these businesses.

As we will see, the German Register of Companies ("Handelsregister") presents a highly attractive candidate for the creation of a sampling frame of newly founded ventures. It not only offers a census-like completeness of founding activity, but the data is also publicly accessible and has been used for the purpose of sampling new business foundations before. One example is the ZEW's data on newly funded enterprises (Almus, Engel, and Prantl 2000; Bretz et al. 2013; Niefert 2008), which prepares data from Creditreform for scientific purposes and therefore relies on CR data (but also other sources). Beyond that, to our knowledge, there are only some smaller studies that have also used the CR for sampling purposes (Braun and Richter 2010; Cantner and Kösters 2012; Dauchert 2013), but none of them have used it for the sampling of entrepreneurial groups.

Step 2: Once we have drawn a probability sample of newly founded business companies, these new business ventures form the starting point from which to identify the set of individuals involved in the particular founding process. When 'new business' is selected in Step 1, a list is created that details all people who hold formal roles in this company. Using this routine, we essentially create a probability sample of all entrepreneurs in Germany where entrepreneurial individuals are clustered within companies. Thus far, the sampling strategy may be compared to a two-stage sampling design for individuals in the wider population, where households are sampled first and individuals second. As we are interested in all individuals engaged in this business, there is no selection of individuals at this stage – all individuals will be considered as potential members of an entrepreneurial group.

Compiling such a list of all individuals occupying formal roles in a business is easier said than done. To do so, we suggest to first check for all shareholders (SH) and managing directors (MD) in the business ventures selected in Step 1. Initially, this list will consist of individuals as well as organizations who, for example, hold shares. As we are only interested in individuals as group members, we need to check again for the people involved in those shareholding ventures and organizations. Depending on the complexity and depth of the shareholder structure, this may need repeating several times. In doing so, we compile a complete list of all individuals engaged in the founding of this particular business, either directly or via their commitment in organizations, who hold shares in the initial, newly-established business. This list forms the basic outline of the entrepreneurial group behind the sampled business venture, but may still encompass further individuals, because all individuals on this list may themselves be engaged in further ventures and organizations.

Figure 1: Identifying Group Members and Business Ventures Related to Entrepreneurial Groups During the Sampling Process



Step 3: As we have outlined in our conceptual considerations, an entrepreneurial group may be involved in several businesses at any one point in time and not all members of a group may be involved in all of the group's ventures at the same time. The list compiled in Step 2 may therefore not entail all members of the group. Rather, it is possible that the group has been involved with other business ventures before and, if we check for individuals holding formal roles in the additional businesses, we may come across potential group members that were not involved in the business venture we have identified at the beginning of the sampling process. To cater to this possibility, we first need to check for additional businesses in which group members currently hold a role as either managing director or shareholder. Second, if we find additional involvements of group members, we essentially need to repeat Step 2 for this additional business, creating a list of all members involved here. All newly identified entrepreneurial individuals should be considered as potential group members as well. The process does not finish here, as these additional individuals may themselves be engaged in even further businesses, which then again need to be checked to see if more individuals qualify for group membership.

From this sketch of a sampling strategy, we can deduce the kind of information a data source must contain in order to be useful in this sampling approach. Figure 1 offers a visualization of Steps 2 and 3 of this strategy. For Step 1, the information needed is a list of all newly established businesses in any one year. For Step 2, information about all entrepreneurial members holding a formal role in this business is required. This information is also

needed for institutional shareholders, until in the end we have arrived at a complete list of all individuals directly or indirectly invested in the organization sampled at Stage 1. For Step 3, the information from Step 2 must allow for the identification of additional involvements of group members in other businesses. For these additional businesses, we once again need to compile the complete list of all persons involved, as all of them are potential members of the underlying entrepreneurial groups. In addition to these already high demands regarding the range of information that is necessary, there is the challenge of generating samples retrospectively. As our intention is to study long-term developments over the last two decades, samples need to be generated that constitute valid cross-sections from previous years. This necessitates that the required information described so far is not (only) available as of the most recent date, but rather at certain time points in the past, e.g., the year 2007. To be able to study changes and developments retrospectively, this will also mean that the data source must contain information about organizations and groups that are no longer active and this information must date back several years at least. Finally, but also importantly, putting together the sampling frame needs to consider the typical time and budget constraints of scientific research projects. Hence, the use of data sources needs to be both cost effective and efficient. This pertains to the ways and formats that data may be browsed and searched, as well as retrieved, exported and formatted for the intended usage. In the next section, we will show how CR data meets the criteria outlined above and may serve as a sampling frame, first for business foundations and founders, then, in a following step, for entrepreneurial groups as such.

3.2 German CR data: Content and Access

The purpose of the CR is to disclose essential information on the legal status of companies in order to improve security and control in business transactions for people dealing with these companies. When taking up commercial activities, entrepreneurs have the legal obligation to register their companies, especially their formation (§§ 29, 1 HGB, "Betriebsbeginnpublizität"), with the German Register of Companies. The registration occurs in their respective district of the registrar, which is generally the local district court ("Amtsgericht", respectively "Registergericht"). The CR is divided into two broader sections. Section A ("HRA") deals with personally run business and sole traders without share capital ("Personengesellschaften").[5] Section B ("HRB") contains all

[5] More specifically, it contains the retail salesman (registered salesman; "eingetragener Kaufmann/ Kauffrau - e.K."), the general partnership ("offene Handelsgesellschaft – oHG"), the limited partnership ("Kommanditgesellschaft – KG"), and the European Economic Interest Grouping (EEIG).

incorporated companies with share capital ("Kapitalgesellschaften").[6] The CR contains information on the legal name of a company, the type of legal entity, its address, the type of its business activity, the people representing the company as either a managing director or an authorized signatory, the names of shareholders, and the amount of share capital, if applicable. Aside from the founding of a business, entrepreneurs also have a duty to report major changes in their data, including any changes to the address, the type of business activity, the top management team, the ownership structure, or cases of insolvency. In addition, companies beyond a certain threshold of annual revenue are required to publish their financial reports. The specific legal requirements regarding the content of the CR are regulated by German trade law ("Handelsgesetzbuch"). This law also lays out the fees that may follow any non-disclosure of required facts and events. Entries to the German Register of Companies are also sent to and published by the "Bundesanzeiger", the official publication of the German government. Starting with fiscal year 2006, annual reports are published in the (electronic) "Bundesanzeiger" and no longer stored with the CR directly.

As CR data is publicly available, several commercial data providers have integrated CR data into their services, which differ in the scope and depth of data offered (Goemann-Singer, Graschi, and Weissenberger 2004). For example, Creditreform is an organization that assesses and sells credit ratings of companies and, as part of its business, has been collecting and storing CR data for a long time. Other information providers, such as Bisnode, Genios, or Bureau von Dijk, do the same (or in fact integrate Creditreform information into their portfolio). Each of these commercial databases comes with particular drawbacks and benefits, which largely depend on the individual circumstances of the project and the specific research question as to which of the sources, or combination of sources, would be beneficial to use. One problem common for several commercial database providers is the effort that they have to make in order to keep all data up to date, which often means that old data is overwritten and lost in the process, or at least not available in their standardized database products. While this makes sense for most of their clients in the private sector, this is problematic for scientific researchers who seek to study longitudinal processes retrospectively. Nevertheless, there are a number of commercial providers that do store historic information – even when this information has been removed from the CR database itself. In addition, the commercial data

[6] These entail the following legal entities: public limited company ("Aktiengesellschaft – AG"), the association limited by shares ("Kommanditgesellschaft auf Aktien – KgaA"), the limited liability company ("Gesellschaft mit beschränkter Haftung - GmbH), the entrepreneurial company ("Unternehmergesellschaft – UG (haftungsbeschränkt"), and other legal forms such as the European public company (SE), the Insurance Society (VVaG), and the pension fund society (PVaG). Other non-German legal formats may also be registered and make up a small fraction of the overall entries.

service providers have undergone the tedious task of making some unstructured and hard-copy materials available online.

3.3 A Proposal to Sample Entrepreneurial Groups from CR Data

In this section, we describe how the proposed sampling strategy for entrepreneurial groups may be implemented by using CR data as the principal data source as well as several other information providers for accessing this data.

Step 1: The proposed sampling approach entails drawing probability samples of newly founded business ventures. For this purpose, we need a full list of all newly founded businesses in one year. Following the sampling strategy outlined above, new entries to the CR may serve as the universe from which we draw our sample. Such new entries are a valid starting point for the sampling procedure, as the vast majority of companies only come into existence by being registered with the CR (in legal terms, the registration is constitutive for the foundation of the limited company. Before that, the limitation of liability is not effective). All new CR entries ("Neueintragungen") are published in the "Bundesanzeiger". There are roughly 100,000 new entries to the register each year.[7] Based on these new entries, it is possible to draw random probability samples, even retrospectively, from previous years.

Data access and longitudinal availability: All new entries to the German CR are published in the Bundesanzeiger, in a print version prior to 2007 and electronically since then. While these new entries are mostly unstructured, making it difficult to extract the required information in an automated fashion, it is basically the raw and original source and hence very valuable. New entries to the CR may be accessed directly through the German CR online portal. However, the electronic online CR only provides access to these texts since its introduction in 2007. Moreover, the online CR is limited in its search options and extraction volume. Fortunately, some commercial data providers (such as Nexis[8]) offer electronic access to these entries in a digitized format, even prior to 2007. Thus, using a commercial provider such as Nexis provides access, in principle, to all new entries to the CR in any given year. However, this access is typically restricted in terms of the amount of information available through these databases in one single step. With Nexis, as it turns out, it is most efficient not to retrieve the full information about new entries, but to retrieve the hit-list(s) of the search on new entries to the CR for a given date. This may

[7] A substantial fraction will not pertain to complete, newly-founded business ventures, but rather to a consequence of moves between court districts of changes in the legal form of a business. These will need to be filtered out as ineligibles.

[8] The Nexis database is a commercial data service that provides large databases of legal and journalistic documents as well as a number of services offering both risk management and computer-assisted legal and business research.

be downloaded and contains the name, date of entry, court district, and (at least since 2007) the CR number of the business that the entry pertains to. This is sufficient information to generate a simple random sample of new business ventures and to match the sample to other data sources in order to enrich the sample with information on individuals or related businesses.

Step 2: This step aims to generate a complete list of those individuals engaged in the business ventures sampled in Step 1. The final list should contain all individual MDs and SHs of these ventures as well as all *individual* MDs and SHs of *institutional* MDs and SHs of the initial ventures. This will be a reiterative process, as there may be several layers of institutional MDs and SHs behind these business ventures.

Data access and longitudinal availability: For this step, we may turn to commercial providers, as the access to additional data, such as the list of shareholders ("Gesellschafterlisten"), is limited. These lists are part of the publicly accessible files on a company (with share capital) and are only provided in the register's appendix. While it is possible to access these lists through the online register (at least for new entries since 2007), they come as pdf documents (or some other picture file format) and cannot be easily processed with statistical programs. Hence, we may rely on commercial data providers (Dafne), which to some extent have already added the content of these lists into their databases. For this process, we enrich the initial information contained in the CR with information taken from commercial databases: the Dafne and Orbis databases, assembled by Bureau van Dijk. These provide central information on the shareholder structure of the selected companies. We require this information to identify all group members (and later to study composition changes and key transitions in the trajectories of the group).[9]

Step 3: The third step is to complete the list of individuals in the groups of entrepreneurs behind the business ventures from Step 1. First, this will require taking the list of individuals from Step 2 and checking for all businesses in which these individuals hold a formal role beyond those identified in Step 1. Second, all these additional businesses need to be checked for individuals not already contained in the list from Step 2. Again, this search will need repeating, as every additional individual may be involved in even more additional businesses, which in turn may entail additional individuals.

Data access and longitudinal availability: The data source must allow searching for business ventures via the individuals who hold a formal position in this venture. Ideally, the source provides an identifier at the personal level,

[9] For the years prior to 2007, these lists mostly existed on paper only and would need to be compiled from court registries scattered around the country before being digitized in some form to be treated as data.

which can be used to search for all ventures that an individual is engaged in (the alternative – to search by name in combination with other attributes such as date of birth – is more error-prone and likely to require substantial manual checking). This exercise may also lead to additional persons engaged in one of the ventures being identified during this step, but not before. Here, it is useful that commercial data providers have set up databases with personal identifiers that allow such "backward" searches of companies via the individuals involved. There are, however, limitations regarding the historic availability of such persons/organizations links, with some sources being richer in histories of managing directors and others being richer in historic shareholder structures, while we ideally require both. Again, a combination of sources for accessment may be needed – and, for samples predating 2007, it may even be necessary to revert to the original public entries from the CR and to do the matching base on the names and dates of birth of managers and shareholders. However, the CR generally entails the information we need to finish Step 3 successfully.

Overall, the German CR appears to be a very useful source of process-generated data for the sampling strategy outlined above. The CR meets the criteria outlined above in the sense that it contains all the information necessary to implement the strategy, going back in time several years. However, the optimal way of accessing CR data and the question of which data provider is to be used will depend on the information required and both the most convenient way to identify the units of interest and the best way to retrieve them, in addition to the data budgetary constraints of the respective research project.

4. An Evaluation of the Quality of German CR Data

Thus far, we have described the strategy for selecting newly founded business ventures and sampling entrepreneurial groups from register data in principle. While we have concluded that the CR is a suitable data source that meets the minimum requirements of a sampling strategy for entrepreneurial groups, we need to reflect on the quality of this administrative data source as a sampling frame. Evaluating potentially systematic errors and lacks of coverage in the CR data is necessary to realize potential threats to the validity and generalizability of analyses conducted with samples drawn from CR data. In the following, we engage in a data exploration and examination exercises with respect to the German CR and reflect upon its consequences for the sample of entrepreneurial groups. For this purpose, we make use of an existing framework for the evaluation of the data quality of process-generated data, of which the CR constitutes an example. We use this framework as a heuristic to identify potential errors and pitfalls in the data, to assess whether these are random or rather systematic, and to what extent they may or may not affect the potential use of the CR as a sampling frame.

Baur (2009) pronounced the relevance of a data source study in using process-generated data in social research. The production of administrative data is dependent on social and institutional practices and is also subject to change over time: "societal and institutional filter influence: 1) which data are produced and how they are produced (production bias), 2) if and how data are stored (selection bias)" (Baur 2009, 12). She pointed to the foundational work of Bick and Müller (1984), which developed a framework to evaluate the data quality of process-generated data collected for administrative purposes in order to reflect critically on the CR as a source of data for scientific inquiry. To this end, Bick and Müller suggest a framework for evaluating process-generated data, which may be used to conduct a source-specific data examination ("Datenkunde") in order to assess the quality of the data and to identify factors that distort data and samples stemming from this source.[10]

Bick and Müller conceived of their scheme to describe and analyze the processes generating so-called social bookkeeping data, i.e., files produced during, among other things, modern governments. Those data collect information on specific parts or cases of interest for the administration, e.g., a person or an object (a house, a car etc., cf. (Bick and Müller 1984)) in such a way that the same type of information is collected in many cases. Due to this quantitative and standardized nature, such process-generated administrative data share many problems with survey data, but they also exhibit some very specific features and problems (Rokkan 1976). In the case of administrative data, the process of data collection and processing is neither preconceived by researchers nor organized around scientific principles of empirical research. Rather, the data collection and preparation process follows guidelines and procedures as they fit purposes that allow public agencies to fulfill their goals and duties (Bick and Müller 1984; Baur 2004). As a result, many factors that possibly affect the data and sampling quality of administrative data are unknown to the researcher. They may well depend on idiosyncratic factors such as individual strategies of handling information of those people responsible for collecting the data. They may also depend on the archival practices and filing systems, which in turn may vary between different government agencies. Consequently, Bick and Müller identify three key elements that determine the content and formatting of such data (Bick and Müller 1984, 128): first, the institutional "logic", i.e., the organizational rules and procedures that govern the collection, storage, and processing of information in the data producing agency; second, the nature of the "client", i.e., the person who is affected by the administrative process and whose cooperation is necessary to produce the data, e.g., by providing

[10] In their view, researchers need to conduct such an evaluation for different data sources separately, as the sources of error and reasons behind them differ too greatly and are too specifically bound to a certain topic, country, or data type as to formulate one all-encompassing error theory of process-generated, administrative data.

information; third, the data agent or administrative "clerk" dealing with the data in the data-producing agency. A fourth important element (and source of potential selectivity in the data) is the archiving process, as someone has to decide if and how data is to be stored and to be made available to other data users, such as scientists (Baur 2004).

4.1 The "Logic" of the Data Producing Agency

Every public agency follows a set of norms and formal procedures for data production which, at least partially, depend on the different uses the agency has for the information they gather. There will be different rules as how to collect, store, and process information, which will also depend on institutional traditions as well as budgetary constraints, the agency network relations and the form and (temporal) structure the agency holds with its clients. All this may well result in inconsistencies between data on the same person from various administrative sources (Baur 2009).

The purpose of the CR is to disclose essential information on the legal status of companies in order to improve security and control in business transactions for people dealing with these companies. Business partners usually have a desire to know about the persons who are legally entitled to speak for a company or to sign contracts before they do business with them. They often also have an interest in knowing about the financial standing of a company, as well as their shareholder structure. The information in the register provides this knowledge and is justiciable, as it may be used in court.[11] The type of information included in the register is determined by this overall purpose of providing the public with knowledge necessary for doing business with companies (rather than some internal administrative logic for fulfilling government functions). There are certain particularities in this administrative process, which may lead to problems with the data entailed in the CR for its use as social science data.

Decentralized organization: While the purpose as well as the content of the register is the same across the country, CR data (particularly prior to 2007) is

[11] The German Register of Companies states on their website: "In a dispute with any third party the company can refer only to such facts which have been duly registered and published or which had already been known to such third party (Section 15 subsection 1 of the German Commercial Code). A third party has to accept such facts that have been duly registered and published. This shall not apply to legal acts which have been taken within fifteen days upon publishing, if the third party is able to prove that it neither had knowledge of the fact nor was required to know such a fact (Section 15 subsection 2 of the German Commercial Code). If a fact had not been registered properly then the third party, in a dispute with the party who is required to register the fact, is entitled to refer to the fact as if it had been published, unless such third party had knowledge of the faulty registration (Section 15 subsection 3 of the German Commercial Code)." <https://www.handelsregister.de/rp_web/legalinfo> (last accessed April 26, 2018).

subject to varying institutional logics on state and district levels. German jurisdiction is organized at the state level, which may result in differences in terms of data handling – one example being minor time differences in the implementation of the electronic transformation. In addition, the maintenance of the German CR is decentralized, which results in different practices and norms of data handling between courts.

Inconsistent identifiers: One example is the fact that courts do not share one logic of assigning IDs to companies (the number of digits is not fixed and some courts add two characters as identifiers for their court district). Consequently, these IDs are not unique across registration courts' districts, which poses a problem when combining data from all areas of the country.[12] Also, as court districts change over time (due to administrative reorganization and downsizing), there may even be changes within the same geographical area. Another consequence of the local, decentralized organization of the original register is the change of location of an established business; if the new location falls into the district of another court district, this will result in a new entry for this business in the new district. As we suggest drawing a sample from the list of new entries per year, these cases represent instances of over-coverage of our sampling frame, as these new entries do not constitute new foundations but mere relocations.

Deleted entries: While the German CR data offers a census-like completeness for new business formations, past information may not arrive at the same level of coverage. It is the purpose of the CR to provide basic information on companies in order to protect their business partners, creditors, and customers. This reason ceases to be valid for companies that have gone out of business. Therefore, entrepreneurs may ask for their entries to be deleted from the register after their business has become inactive ("Löschungen"). After the deletion, one cannot retrieve information on these companies any longer. This may be especially problematic if those entrepreneurs who do ask for their entries to be deleted from the register differ from those who do not make this active request. However, it is still possible to check the CR public entries from the past for these companies, as these entries are archived in commercial databases. Some of the missing information may be restored this way, while others, unfortunately the list of shareholders among them, are likely to be lost.

Changes in the legal framework: In a longitudinal perspective, it is also important to check for legal changes regarding the registration procedure that may have occurred over time. For example, in 2008, a new, simplified procedure to

[12] Luckily, there is a relative easy fix for this problem: by adding the postcode of the courts district to the ID they should become unique. Another problem, however, may be the longitudinal uniqueness of IDs. This problem may arise when an ID previously used for one company, that for example has gone bankrupt, is later assigned to a new company. Whether such problems exist probably also depends on the practice within each court district.

register GmbHs was introduced to ease the burden of registration as long as certain criteria were met (§2 Absatz 1a, GmbHG).

4.2 The "Clerk" and the Data Formatting Process

The agents dealing with the information in the administrative institution, i.e., the data-producing clerks, are an important factor when considering the quality of process-generated data. How they perform their task will largely depend on a range of personal factors, their education and training in handling and processing data, their general abilities and attitudes towards clients, their "customer orientation", stamina, work ethic, adherence to rules and attention to detail, their role description, performance expectations, and remuneration within the specific organization (Baur 2004). It may also depend on their relation with other agents performing the same tasks or similar tasks at other steps of the "production line" of the data.

In the case of the German CR, data formatting used to be the responsibility of administrative data clerks, but the processing procedure has changed over time and has moved beyond the manual editing of data. At the beginning of 2007, the register experienced a dramatic change, with the transition into a strictly electronic format. In fact, the data processing clerk has become largely absent from the data generating process because the "clients" now have to provide their information themselves, through specifically programmed electronic and web-based routines. This renders the manual work of handling data by clerks mostly redundant – a development Bick and Müller may not have been able to foresee. Hence, rather than just referring to the level of "data producing clerk," we speak more broadly of the data formatting process. This change also has many possible implications for data quality and the automatized handling of the CR data, especially in a longitudinal perspective.

Decentralized administration/diverging file systems: The German CR used to be organized locally by the respective district courts, of which there are about 110 in the country. All of those administrative bodies adhered to traditional ways of manually maintaining flashcards and paper files. While the content of the files was standardized for the whole country, the way files were kept and stored may well have been different between court districts. Potentially, the possibility to retrieve information from the years prior to 2007 may differ greatly as well.

Timeliness of reporting/Random Errors (Typos): In general, by the introduction of the electronic register, the scope for individualized error types by data-producing clerks had diminished greatly. Also, the time that passes between the entry into the register, the date of registration, and the date of its publication is likely to be much smaller than before, when all this had to be registered on paper first and then sent to the Bundesanzeiger for publication. However, the clients themselves probably now commit the type of errors that had previously

been committed by the data clerk. An illustrative example is the new legal form of small limited companies ("Unternehmergesellschaft"). These need to carry the extension "UG (haftungsbeschränkt)" in their name. In the list of the small, limited companies of 2012, there are probably a dozen different misspellings of the word "haftungsbeschränkt". This exemplifies the scope for random errors that still exist, despite the electronic administration of the CR, and may have consequences for the identification of companies across register entries if not checked and corrected. At the same time, typically for centralized structures, the flexibility is likely to be reduced and if errors occur, they often affect the whole system at once. Still, this centralized procedure is probably less error-prone overall, as this set-up bypasses the type of errors that are exactly the focus of Bick and Müller's discussion of the data clerk. However, instead of the idiosyncratic characteristics of data clerks, in terms of data quality, one now has to consider the specificities of the electronic system set up for the information transfer. This routine determines to a large extent the structure and content of the data source, with all the advantages and disadvantages of a centralized procedure.

4.3 The "Client" Providing the Information

The clients of administrative agencies themselves are important, as they may also influence the process of data production, depending on their ability and willingness to provide and reveal possibly personal information. The clients' willingness may be shaped by their acceptance of administrative rules, the agency's interference with the life of their clients, and the rewards and drawbacks that clients face if they cooperate or do not cooperate with the administrative agency. Other factors may be the medium and timing of contact with the agency, as well as their involvement in other realms of life, such as their families and the workplace. In extreme cases, clients may even forge data (cf. Lippe 1998, Salheiser 2009), as they may deem some information to be irrelevant, or simply forget to tell the data-producing agency.

A particularity of the German CR is that the client-administration relationship is less pronounced than in other cases of administrative process-generated data. While the "client" is the business about which information is kept in a file at the CR, the beneficiary of that information is neither the business (i.e., the client) nor the CR (i.e., the administration), but third parties inquiring about companies they intend to do business with. Still, the German CR relies strongly on the companies to provide data and information, a dependency that has even increased with the implementation of the electronic register.

Non-reporting: Although there is a legal obligation to report information to the CR (see above) and there is a risk of being sued based on misleading public information when failing to report, anecdotal evidence suggests that entrepreneurs do not report every change in the business structure to the CR. In particu-

lar, the need to notarize entries to the register – such as the establishment of a new business – as well as any changes to the address, the type of business activity, the management, or cases of insolvency implies the necessity of time, effort, and legal fees for reporting changes to the register and hence reduces the willingness to do so. In addition, the authorities did not follow up on the legal publication obligation for a long time. It has only been since 2007 that the fees have been implemented and authorities have traced overdue companies (Fockenbrock and Fröndhoff, March 25, 2008). Thus, the information in the register will depend on the willingness of companies to disclose information, especially business sensitive data, even though it may be required by law. Finally, firms may shy away from publishing too much information on their business, as they may see it as a competitive disadvantage (in particular with regard to annual financial statements). There is anecdotal evidence that some firms are quite willing to pay resulting penalty fees rather than publish financial information for everyone to see. As the fees do not exceed a certain maximum, larger corporations especially may opt to "buy their way out" of publication obligations. These various reasons may lead to the underestimation of change in group composition and missing data, especially for young companies. In addition, financial information before 2008 will be missing for many companies, which may constrain the possibility of longitudinal comparisons.

Non-standardization: Another important point in terms of data quality is that not all information has to be submitted to the German CR in a standardized manner. Take, for example, the provision of a list of shareholders: this information can be provided in an electronic document (such as Word), notarized and scanned if necessary, and sent to the register in an electronic format such as a pdf or a picture file. While there is regulation on which information is required, the actual format of the notification varies widely, based on personal taste and competencies. Both the format as well as the unstandardized content poses strong limitations on the possibility to process these lists of shareholders automatically. However, commercial databases offer this information (based on CR data) in an electronic and standardized format, at least since these documents have been available as electronic copies from the registers.

Timeliness of reporting: Even if all companies provide all legally requested information and changes (and do not differ in their habit to demand a deletion if applicable), there may still be differences in the timeliness of their provision of information to the register; some companies may require longer to report a change in the shareholder structure than others. Hence, the accuracy of the timing of events may vary between companies if they have to report these themselves.

4.4 Data Archiving, Release and Retrieval

Beyond the three points discussed above, there are factors affecting the sample quality of process-generated data, as Baur (2009) points out, which relate to the practices of archiving those data and the ways in which they are released to the public and finally accessed by researchers. In this process, the technology of data storage and processing by computers for further analysis marks a key element. The usefulness of the data for longitudinal studies depends particularly on the availability of past data and the rules for archiving files and information.

Online access: An easy and efficient way to access the register is via the website of the joint register portal of the German federal states.[13] The online register allows searching for individual firms based on their name and court district; the register ID is not necessary. The most basic information on each business ("Unternehmensträgerdaten – UT") as well as simple publications ("VÖ – Veröffentlichungen") can be accessed free of charge. It is also possible to retrieve pdf printouts (formerly known as "Handelsregisterauszug" or "HR-Auszug"), but a fee will be charged and prior registration is necessary.[14] One printout provides all currently valid entries while the so-called chronological printout contains all information since 2007, the year of the electronic transformation. Finally, the so-called "historic" printout contains all information on a business from when the register was still kept on paper before 2007. In addition, several documents are available that belong to special volumes or appendices, as long as they have been transmitted electronically to the district court. This may entail the application to the CR as well as, for example, the lists of shareholders and founding documents of limited companies and the minutes of shareholder meetings.

Regulation of use: According to the register's purpose of providing the public with basic information on companies, in principle, anybody may request and extract information about a specific business. However, there are several ways to do so and they differ in terms of access, practicability, and cost. In the electronic online version of the CR, for example, one is bound by copyright legislation[15] as well as the regulations of use, which permits "creating or maintaining one's own register as a counterpart to the Register of Companies and from providing the data retrieved by anyone for such purposes." In addition, one must "limit the frequency of one's requests to 60 per hour", which limits the possibility of automated web-scraping of the data.[16] The website also offers

[13] <www.handelsregister.de>.

[14] As of March 2018, the fee was €4.50 per printout and company.

[15] "If not indicated otherwise, all pages on this information server are protected by the copyright law." <https://www.handelsregister.de/rp_web/legal-info.do>.

[16] <https://www.handelsregister.de/rp_web/nutzungsordnung.do>. However, the regulation of use also allows additional access time for a limited period: "if the user is able to prove that

some meta-data available that could be very valuable, depending on the research question. Examples of this are the dates when some documents, such as lists of shareholders, were received by the CR, as this meta-data might be a good indicator of changes in shareholder structure as well as the date they actually happened, which is difficult if not impossible to get from other sources.

Formatting and availability: There are also limits to what is provided and what is available, especially over time. Once companies have gone out of business, their register entry has served its legal purpose. As a common practice, old files of inactive companies remain in the register for ten years. What happens afterwards seems to depend on local court practice: files may be stored in a public archive or may even be destroyed. Even if files are kept in an archive, retrieving them is likely to be costly and burdensome. Also, the above-mentioned deletions ("Löschungen"), which may be requested by business owners, limit the availability of data retrospectively, although it may be possible to regenerate some of the data from entries to the Bundesanzeiger ("Bekanntmachungen"), rather than the actual register files on a business, as these public entries cannot be deleted. Certain information, however, especially documents in the "appendix" to the file such as the lists of stakeholders, will be lost. For these and other instances, it is useful to check the availability with commercial data providers who still have this information on file. Two prominent examples are Bisnode (previously "Hoppenstedt Firmeninformationen") and Bureau van Dijk, which typically sell timely limited access to their databases, usually for a one-year minimum. In the database, you can search for specific companies, compile lists and reports, and download data to a certain extent. Some types of information are more restricted than others, such as financial and contact information, which may be used for direct marketing purposes. Bureau von Dijk offers at least three different databases: Markus, a database of publically registered companies in Germany; Dafne, a database of companies that compile annual financial statements; and Orbis, a database containing the former two as well as information on current and previous companies worldwide. While Markus covers more companies, Dafne offers more information per business. Regarding content, contract conditions, and handling of the web-based database interface, Bisnode and Bureau van Dijk are relatively similar. However, coverage in the Bureau van Dijk databases seems much broader and more census-like, which is a problem for the Bisnode database. Another possibility to build a dataset is to use so-called data curation services, which enrich and/or update existing databases from customers with information from their own databases whenever fields of information or whole cases are missing.

---

he has a justified reason for accessing the data more frequently" and only "upon prior consultation with the Register Portal service department."

5. Potential Coverage Issues in the CR Data

After having discussed potential threats to data quality in the CR due to different sources, we now consider how these errors may affect the multi-step sampling approach that we have outlined above. We address each step of the sampling process, i.e., each level of units involved, separately: the level of newly founded business first, the level of entrepreneurial individuals second, and entrepreneurial groups third. In the following, we will specifically discuss problems of representation in the CR data (Groves et al. 2011), including measurement errors that lead to such representation problems.[17]

Kish (1995) considers four basic problems of sampling frames: (1) missing elements – elements of the population that are not included in the frame; (2) foreign elements – elements included in the frame that do not actually belong to the population; (3) duplicate entries – elements that appear more than once in the frame; and (4) clusters of elements – the sampling frame entails clustered sampling units instead of individual elements. In other words, a precondition for successful representation is a sampling frame that matches the population, i.e., ideally exhibits no coverage error and thus contains *all* elements of interest in the population and *only* such elements of interest (Groves et al. 2011). However, in most cases, sampling frames will be less than perfect and will exhibit some form of coverage error. Over-coverage relates to the fact that there are more elements in the list than those we actually need (Kish's first problem), while under-coverage describes the fact that elements are missing (Kish's second problem). Over-coverage is not so much of a problem, as these cases are actually not part of the population under study (usually called *ineligible* for the purpose of the study) and hence can be removed from the sample after the sample has been drawn. Ineligibles, however, do pose three problems: first, their identification may require time and incur cost; second, they reduce the final sample size (this is especially a problem if the research design does not allow the re-sampling of cases from the frame); and third, the presence of ineligibles in the sampling frame makes the calculation of sampling probabilities for the actual units of interest more difficult. Still, under-coverage is more problematic because certain elements of the population have zero chance of being selected for the study. This will be an even greater problem if those missing cases systematically differ from the remaining cases in the list (which is most definitely the case for homeless people, even though their number is relatively small compared to all German nationals overall). The other two of

[17] For example, the misspelling of a name may prevent a person or a company from being identified in the database and hence may not become part of the sample, even though they should be. It also makes the matching of people across different entries with various businesses more error-prone.

Kish's problems are quite unproblematic, at least at the first step of the sampling strategy: duplicates may simply be identified by means of the name of the company, the CR ID, and the date of entry. Clusters of companies are not to be expected as the CR defines individual organizations as new entries. However, clusters do appear at the second step, where individuals are clustered within organizations. This, however, does not pose a problem but rather is part of the sampling design.

Regarding over-coverage and under-coverage, it is important to think about the consequences that these coverage errors may have, especially in terms of their systematic nature and potential threat to the analysis of the data and the dependent variables used. What are the potential biases of CR data for our study of the long-term development of entrepreneurial groups? Following our conceptual distinction between business organizations, entrepreneurial individuals, and entrepreneurial groups, it is necessary to consider each of these levels separately when evaluating coverage issues of the drawn samples.

5.1 The Coverage of New Business Ventures

In the German CR, both over-coverage and under-coverage of companies is an issue. First, over-coverage appears in the data, based on ineligible new entries to the CR, which needs to be removed from the sampling frame, or at least from the resulting sample. In our case, ineligibles may consist of companies run by individual entrepreneurs ("*eingetragene Kaufleute*"), as they do not qualify as an entrepreneurial group as per our definition. The same holds true for other legal entities that do not constitute economic businesses, such as voluntary associations, partnerships among the free professions, cooperatives, and state-run companies. Charitable companies ("gGmbH", "gUG (*haftungsbeschränkt*")) constitute a different case. Finally, relocated companies or newly established branches of foreign companies are not considered as new foundations and are hence regarded as ineligibles and should be removed.

A difficult case the so-called shelf companies ("*Vorratsgesellschaften*"), are a business model whereby a firm sets up new, "empty" firms that it sells on to entrepreneurs who quickly want to start a business without having to go through the bureaucratic steps of implementing and registering a company themselves. The process of buying such a shelf company, changing its name, and shifting its business purpose is still quicker and less burdensome than setting up everything yourself (these changes will again need to be registered with the CR). This leaves us with the problem that sometimes several hundreds of these shelf companies are registered at once but are not new businesses in the strict sense. They only become so after they get bought by the new owners, who are the "real" entrepreneurs. The shelf companies are thus of no interest for our research, yet leaving them out will mean that we miss out on those

entrepreneurs who found their business by buying and re-registering such a shelf company.

Under-coverage pertains to those companies that could not be sampled but should be in the sample. The completeness of the German CR is limited by its very constitutive requirement – only limited companies or those companies that exceed a certain level of economic activity have the obligation to register. This definitional criterion largely underestimates owners of micro-businesses with low economic activity. However, these are almost by definition individual entrepreneurs and hence do not fall within the scope of our research.

5.2 The Coverage of Founders of New Businesses

In Step 2, we have to ask to what extent the data sources used contain a full and correct picture of the managing directors and personal shareholders involved in a given business. Again, in principle, all individuals holding a formal role in the business should be identifiable with CR data. In 2007, a big decisive event was when the CR became a purely electronic register. The information contained in the CR has been integrated by many commercial information brokers into their databases since then. Usually, it is relatively easy to generate a list of all managing directors and shareholders of any given business, regardless of the data provider. This is, however, mostly true for the current set-up of the business. If there have been changes regarding directors or shareholders, historic information is needed and, here, data providers differ between what they can offer and for which time frame. While some only offer a complete picture of past and present shareholders, others provide this for executives only, at least dating back to 2007. This may affect the comparability of the samples over time. Also, both the information on the shareholders, as well as on managing directors and executives (when changes occur), depend on the clients' behavior and their cooperation with the CR to provide this information accurately and on time. If there are stark differences between rule abiding and non-rule abiding organizations (that either provide legally requested data late or not at all), this may lead to missing data regarding, for example, the list of shareholders of a business. Despite the reduction in sample size, such absence may be problematic if it is related to some important underlying characteristics of the companies or groups. The original list of managing directors should, however, always be available because this is part of the new entry that the organizations are sampled from.

5.3 The Coverage of Entrepreneurial Groups

Given our approach for the identification of entrepreneurial groups, it is important to consider what kind of groups will be over-sampled or under-sampled and how systematic these errors will be. Some things that we will not be able to cover, given the nature of our data source, are the very early stages of group

formation. At this early point, group ties may be very loose, as people vaguely develop an idea together, but it is still unclear who will actually devote time and effort to set up a joint entrepreneurial project. Some people will drop out while others will join before the business is registered and we will not be able to study this potentially very dynamic stage with our empirical approach.

This also relates to the fact that the German CR data only allows the study of certain formal roles. It is only possible to identify chief executives and shareholders as group members. This way, we will not be able to investigate other persons who might be considered team members under a different operationalization of group membership, maybe with less formal supporting roles (e.g., in a primarily advisory capacity). Instead, we have to rely on membership as defined by formal and institutionalized legal role definitions. By leaving out potential, informal, and self-identified group members, we may miss out on developments and activities which are potentially faster and more volatile than the coming and going of formal group members.

Transnational group members and border-crossing business activities pose a different kind of problem. As we rely on data from a German registry, we do not know if a group sets up another business venture in another country. Similarly, commercial databases, depending on the type of access you purchase, are likely to keep information on German residents and firms only. This means that we are likely to only identify group members and businesses with residence in Germany. This in turn may lead to some selectivity in the sample, as border-crossing activities are more likely in some industries than in others.

6. Discussion

In this article, we proposed a sampling strategy for the creation of a representative sample of entrepreneurial groups in Germany. From our conceptual discussion in conjunction with our research goal to study entrepreneurial groups longitudinally and retrospectively, we derived a strategy by which we can sample entrepreneurial groups using a procedure akin to hypernetwork sampling. This was necessary for two reasons: previous research has mainly used (1) non-representative samples that contributed to the inconclusiveness of empirical results; and (2) inadequate conceptual depictions of entrepreneurial groups. With our proposal, we seek to remedy both of these shortcomings. Most importantly, we see entrepreneurial groups and the businesses they are engaged in as separate entities that are related to one another in complex ways.

In sum, the proposed sampling strategy consists of a three-step procedure: (1) identifying newly founded businesses; (2) extracting all individuals involved in entrepreneurial activity; and (3) finally searching for additional businesses that these individuals are engaged in, including their business partners. This procedure yields a complete list of potential group members and their

business(es) in a given year. This task is straightforward as long as an entrepreneurial group is attached to only one business or all group members are active in all businesses. However, depending on the entrepreneurial activity of group members and the shareholder structures of the companies involved, the identification of groups may become complicated. On this premise, it may become necessary to calibrate the boundaries of an entrepreneurial group, i.e., to narrow down the decision of whom and what businesses belong to the entrepreneurial group. In certain cases, e.g., where a group is associated with multiple businesses and where members play different roles across these ventures, group composition may become overly complex. In order to reduce complexity, two criteria seem significant: the type of co-engagement in a business (e.g., owner and manager, co-owners, investors) and a minimum number or percentage of co-present group members across businesses. It may become necessary to use cut-off points on these two criteria to prevent groups from becoming too large. Also, it is a possibility to exclude entrepreneurial individuals as group members if they are only linked by investment ties. However, the extent of this problem is an open empirical question and practical solutions will depend on the specific research questions which guide the study.

As we have seen, this sampling approach has certain requirements for a data source to be used as the basis for sampling. Overall, the CR meets the requirements well. The use of the German CR for sampling entrepreneurial groups offers an exemplary solution to the pressing methodological challenge of drawing valid samples of entrepreneurial groups retrospectively, a necessary step before harnessing the opportunities of learning about such groups. The CR is especially appealing for generating a longitudinal dataset of entrepreneurial groups using process-generated, administrative data – something hardly possible with survey data. Using the CR may enable their trajectories to be studied by following events that occur along their group development, such as attachment and dissolution. The discussion of the data quality of the CR has shown that while there may be errors, for example due to the data processing by the local district courts or due to data provision by the businesses, these are likely to be non-systematic and do not impact the overall quality of the sample. However, some forms of under-coverage are likely to exist on all three levels – the business and the individuals as well as the groups. This is most likely a problem in the longitudinal dimension and one should therefore be careful to draw a comparison over time.

6.1 Limitations of our Proposal

While our overall assessment of this data source is positive, there are certain limitations that relate to the conceptual decisions we have made regarding the definition of group boundaries and the sampling strategy we propose in order to study group trajectories over time. Most profoundly, the use of administrative

process-generated data for the sampling of groups restricts our observations to formal roles such as executive management and ownership. Therefore, there are certain research questions that, by definition, may not be addressed using such a research design.

First, as we proposed to define group membership by occupying a formalized role within at least one of the respective businesses, this definition is somewhat restrictive and does not mirror additional forms of engagement. As mentioned earlier, this definition excludes all informal ways of engaging in an entrepreneurial project, for example, through advice, labor, or emotional support. Nevertheless, this operationalization allows the identifying of a set of individuals collectively engaged in more formal roles in a business that can now be tracked for their engagement in additional entrepreneurial activities over time.

Second, some things that we will not be able to study using CR data are the types of tie that binds group members together (e.g., family members or friends) and how these ties might change over time. In addition, CR data contains little information on the internal role differentiation of entrepreneurial groups, such as who the leader of the group is (if there is one) or who the creator and inventive spirit is. For example, internal structures may be democratic or authoritative, discussions may be vivid or inert, and decision processes may result in conflict or consent. For these kinds of internal divisions and dynamic relationships among group members, we would need to augment the sample we have with data from different data sources, such as data from social networking sites, information taken from business websites, or newspaper articles written about the entrepreneurial group and its members.

Third, we will not be able to study the early formation phase of entrepreneurial groups, as we sample only groups who have already established themselves. Hence, we also exclude cases where an entrepreneur starts individually and takes on partners and investors later on. However, if one wanted to study such processes, one would likely need to use other data sources, such as a sampling frame (such as the "Gewerbeanzeigenstatistik"), because many of the individual owner-led small-business companies are not required to register with the CR and are not included there.

6.2 Outlook: The Longitudinal Study of Group Trajectories

Overall, the sampling strategy presented here is most promising for the longitudinal but retrospective analysis of the evolvement of entrepreneurial groups over time using process-generated data. From such a longitudinal perspective, it makes sense not only to understand entrepreneurial groups as a distinct analytical unit, but also to capture their pathways. Such an analysis will, however, require further conceptual work on the evolvement of entrepreneurial groups

over time and the use of process-generated data that contains the information necessary to capture their trajectories.

The conceptual work would likely need to entail a typology of those events potentially involved in the evolution of an entrepreneurial group as well as a conceptual framework for capturing the overarching structure of these events and how they connect with each other. Borrowing from life course research (Elder 1985; Elder, Kirkpatrick Johnson, and Crosnoe 2003), we tentatively suggest to use the term "trajectory" to describe the path travelled by entrepreneurial groups from their emergence to their dissolution. Within these trajectories, a number of transitional events occur that mark the end of one sequence of activities and the beginning of another. Such a longitudinal perspective on entrepreneurial groups has the opportunity to integrate research on the formation, development, and dissolution of different types of entrepreneurial groups. It enables the tracing of dynamic groups in relation to the life courses of the engaged members and in co-evolution of the respective organizations. However, this theoretical framework would require further refinement before it can produce testable hypotheses for empirical research.

Empirically, by tracking their collective engagement over time, we could learn about the stability of the group, new member entries or exits, and additional businesses they engage in. This would require, as a bare minimum, checking whether additional persons have left or joined any of the businesses associated with the entrepreneurial group, thereby either leaving the group or potentially becoming new group members. It will be the pathway of our research group to continue our research in this line, to work on these theoretical issues, and to set up a longitudinal database with which we can study the evolution of entrepreneurial groups over time. By offering a strategy for the sampling of entrepreneurial groups retrospectively, based on process-generated data, we hope other researchers join our scientific endeavor so that together we can work towards a valid and representative understanding of these phenomena.

7. Conclusion

In this paper, we set out with the observation that previous research was deficient in two important but interrelated ways. First, the complex nature of entrepreneurial groups, which are distinct from the businesses that individual group members are engaged in and which may evolve significantly over time, had conceptually not been adequately captured. Second, empirically, existing studies used mostly small, specialized, non-representative samples that did not allow the drawing of valid conclusions on the broader phenomena of entrepreneurial groups. In this paper, we took a first step towards mitigating this unsatisfactory state of research by providing a more complex conceptual understand-

ing of entrepreneurial groups and by putting forward a sampling strategy in order to arrive at proper samples on which valid research can be based, especially in the longitudinal dimension.

We also discussed to what extent the German CR meets all the necessary requirements to serve as the right data source for implementing this sampling strategy. Given the information entailed in the CR, it is possible to identify group members via their role in registered businesses (and other organizations) successfully. Our discussion has shown that CR data is a useful sampling frame that allows for the identifying of a set of individuals collectively engaged in a business and tracking this set of individuals in terms of their engagement in additional entrepreneurial activities across time. Finally, we also discussed certain pitfalls of using this data in terms of over- and under-coverage of certain types of organizations and individuals. Depending on the focus of the analysis, these should be remembered when discussing empirical findings and results, especially for comparisons over long periods of time.

References

Almus, Matthias, Dirk Engel, and Susanne Prantl. 2000. *The Mannheim Foundation Panels of the Centre for European Economic Research (ZEW).* ZEW-Dokumentation 00-02. Mannheim: ZEW.

Baur, Nina. 2004. *Wo liegen die Grenzen quantitativer Längsschnittanalysen?* Bamberger Beiträge zur empirischen Sozialforschung 23. Bamberg: Otto-Friedrich-Universität Bamberg.

Baur, Nina. 2009. Measurement and selection bias in longitudinal data. A framework for re-opening the discussion on data quality and Generalizability of Social Bookkeeping Data. *Historical Social Research* 34 (3): 9-50. doi: 10.12759/hsr.34.2009.3.9-50.

Berrone, Pascual, Cristina Cruz, and Luis R. Gomez-Mejia. 2012. Socioemotional wealth in family firms: Theoretical dimensions, assessment approaches, and agenda for future research. *Family Business Research* 25 (3): 258-79.

Bick, Wolfgang, and Paul J. Müller. 1984. Sozialwissenschaftliche Datenkunde für prozeßproduzierte Daten: Entstehungsbedingungen und Indikatorenqualität. In *Sozialforschung und Verwaltungsdaten*, ed. Wolfgang Bick, 123-59. Stuttgart: Klett-Cotta.

Block, Jörn, Lars Hornuf, and Alexandra Moritz. 2018. Which updates during an equity crowdfunding campaign increase crowd participation? *Small Business Economics* 50 (1): 3-27.

Bonardo, Damiano, Stefano Paleari, and Silvio Vismara. 2010. The M&A dynamics of European science-based entrepreneurial firms. *The Journal of Technology Transfer* 35 (1): 141-80.

Braun, Susanne, and Jörg Richter. 2010. *Gründungsaktivitäten im Handelsregisterbezirk Lüneburg: Eine empirische Studie zur Unternehmergesellschaft.* Lüneburger Beiträge zur Gründungsforschung, 8.

Bretz, Michael, Georg Metzger, Margarita Tchouvakhina, Katrin Ullrich, Jürgen Egeln, Sandra Gottschalk, Martin Murmann, and Michaela Niefert. 2013. *KfW/ZEW-Gründungspanel 2013: Junge Unternehmen trumpfen auf*. Research Report. Mannheim: Zentrum für Europäische Wirtschaftsforschung (ZEW).

Breugst, Nicola, and Dean A. Shepherd. 2017. If you fight with me, i'll get mad! A social model of entrepreneurial affect. *Entrepreneurship Theory and Practice* 41 (3): 379-418.

Brinckmann, Jan, and Martin Hoegl. 2011. Effects of initial teamwork capability and initial relational capability on the development of new technology-based firms. *Strategic Entrepreneurship Journal* 5 (1): 37-57.

Brüderl, Josef, Peter Preisendörfer, and Rolf Ziegler. 1992. Survival chances of newly founded business organizations. *American Sociological Review* 57 (2): 227-42.

Cantner, Uwe, and Sarah Kösters. 2012. Picking the winner? Empirical evidence on the targeting of R&D subsidies to start-ups. *Small Bus Econ* 39 (4): 921-36.

Claessens, Dieter. 1983. Die Gruppe unter innerem und äußerem Organisationsdruck. In *Gruppensoziologie: Perspektiven und Materialien: Perspektiven und Materialien*, ed. Friedhelm Neidhardt. Special issue, *Kölner Zeitschrift für Soziologie und Sozialpsychologie Sonderhefte (*25): 484-96.

Cooney, Thomas. 2005. What is an entrepreneurial team? *International Small Business Journal* 23 (3): 226-35.

Dauchert, Helge. 2013. *Internet- und IT-Unternehmensgründungen in Berlin: Ursachen und Perspektiven des IT- und Internet-Gründungsbooms.* Studien zum deutschen Innovationssystem 13-2013. Berlin.

Elder, Glen H. 1985. Perspectives on the life course. In *Life course dynamics: Trajectories and transitions, 1968-1980*, ed. Glen H. Elder, 21-49. Ithaca: Cornell University Press.

Elder, Glen H., Monica Kirkpatrick Johnson, and Robert Crosnoe. 2003. The emergence and development of life course theory. In *Handbook of the life course*, ed. Jeylan T. Mortimer and Michael J. Shanahan, 3-22. New York: Kluwer Academic.

Fockenbrock, Dieter, and Bert Fröndhoff. 2008. Viele Firmen ignorieren Publizitätspflicht. *Handelsblatt*, March 25, <www.handelsblatt.com/unternehmen/mittelstand/jahresabschluesse-viele-firmen-ignorieren-publizitaetspflicht/2938108.html> (Accessed April 26, 2018).

Goemann-Singer, Alja, Petra Graschi, and Rita Weissenberger. 2004. *Recherche-handbuch Wirtschaftsinformationen: Vorgehen, Quellen und Praxisbeispiele; mit 17 Tabellen,* 2nd ed. Berlin: Springer.

Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2011. *Survey methodology,* 2nd ed. Hoboken: John Wiley & Sons.

Harper, David A. 2008. Towards a theory of entrepreneurial teams. *Journal of Business Venturing* 23 (6): 613-26.

Hartmann, Josef. 2017. Stichprobenziehung und Feldzugang in Organisationsstudien. In *Handbuch Empirische Organisationsforschung*, vol. 4, ed. Stefan Liebig, Wenzel Matiaske, and Sophie Rosenbohm, 185-211. Wiesbaden: Springer Fachmedien Wiesbaden.

Homans, George Caspar. 1972. *Theorie der sozialen Gruppe,* 6th ed. Opladen: Westdeutscher Verlag.
Johannisson, Bengt. 2014. The practice approach and interactive research in entrepreneurship and small-scale venturing. In *Handbook of Research Methods and Applications in Entrepreneurship and Small Business*, ed. A. Carsrud and M. Brännback, 228-58. Cheltenham: Edward Elgar.
Kalleberg, Arne L. 1996. *Organizations in america: Analyzing their structures and human resource practices.* Thousand Oaks, Calif.: Sage Publications.
Kamm, Judith B, Jeffrey C. Shuman, John A. Seeger, and Aaron J. Nurick. 1990. Entrepreneurial teams in new venture creation: a research agenda. *Entrepreneurship Theory & Practice* 14 (4): 7-17.
Kish, Leslie. 1995. *Survey Sampling.* Wiley Classics Library. NY: Wiley.
Klingbeil, Caren, and Thorsten Semrau. 2016. For whom size matters – the interplay between incubator size, tenant characteristics and tenant growth. *Industry and Innovation* 24 (7): 735-52.
Klotz, Anthony C., Keith M. Hmieleski, Bret H. Bradley, and Lowell B. Busenitz. 2014. New venture teams: a review of the literature and roadmap for future research. *Journal of Management* 40 (1): 226-55.
Kmec, Julie A. 2003. Collecting and using employer-worker matched data. *Sociological Focus* 36 (1): 81-95.
Koeller, C. Timothy, and Thomas G. Lechler. 2006. Economic and Managerial Perspectives on New Venture Growth: An Integrated Analysis. *Small Business Economics* 26 (5): 427-37.
Lechler, Thomas. 2001. Social interaction: A determinant of entrepreneurial team venture success. *Small Business Economics* 16 (4): 263-78.
Leiß, Gabriela. 2014. *Ich seh das Ganze als Staffellauf: Intergenerative Aushandlung im Nachfolgeprozess von Unternehmerfamilien.* Heidelberg: Carl Auer Verlag.
Marquis, Christopher, and András Tilcsik. 2013. Imprinting: toward a multilevel theory. *Academy of Management Annals* 7 (1): 195-245.
Neidhardt, Friedhelm, ed. 1983. Gruppensoziologie: Perspektiven und Materialien Special issue, *Kölner Zeitschrift für Soziologie und Sozialpsychologie Sonderhefte* (25).
Niefert, Michaela. 2008. Das neue KfW/ZEW-Gründungspanel: Ergebnisse der ersten Befragungswelle. *ZEW Gründungsreport* 8 (2): 3-4.
Nordstrom, Onnolee A., and Jennifer E. Jennings. 2015. Charting the collective interest in collective entrepreneurship: An integrative review. *Academy of Management Proceedings* 2015 (1): 16851.
Preller, Rebecca, Nicola Breugst, and Holger Patzelt. 2015. Do we all see the same future? The impact of entrepreneurial team members' vision on team and venture development. *Frontiers of Entrepreneurship Research* 35 (8): 196-201.
Rokkan, Stein. 1976. Data Services in Western Europe. Reflections and Variations in the Conditions of Academic Institution Building. *Amercian Behavioural Scientist* 19: 443-54.
Ruef, Martin. 2010. *The entrepreneurial group: Social identities, relations, and collective action.* Kauffman Foundation series on innovation and entrepreneurship. Princeton, NJ: Princeton University Press.

Ruef, Martin, Howard E. Aldrich, and N. M. Carter. 2003. The structure of founding teams: Homophily, strong ties, and isolation among U.S. entrepreneurs. *American Sociological Review* 68 (2): 195-222.
Schäfers, Bernhard, ed. 1994. *Einführung in die Gruppensoziologie: Geschichte, Theorien, Analysen,* 2nd ed. Heidelberg, Wiesbaden: Quelle und Mayer.
Schäfers, Bernhard. 1999. Entwicklung und Grundlegung der Gruppensoziologie. In *Einführung in die Gruppensoziologie: Geschichte, Theorien, Analysen*, ed. Bernhard Schäfers, 3rd ed., 19-36. Wiesbaden: Quelle und Meyer.
Schjoedt, Leon, and Sascha Kraus. 2009. Entrepreneurial teams: definition and performance factors. *Management Research News* 32 (6): 513-24.
Schjoedt, Leon, Erik Monsen, Allison Pearson, Tim Barnett, and James J. Chrisman. 2013. New venture and family business teams: Understanding team formation, composition, behaviors, and performance. *Entrepreneurship Theory & Practice* 37 (1): 1-15.
Schnell, Rainer, Elke Esser, Paul B. Hill. 2005. *Methoden der empirischen Sozialforschung,* 7th ed. München: Oldenbourg.
Schumpeter, Joseph A. 1965. Economic theory and entrepreneurial history. In *Explorations in Enterprise*, ed. H.G.J. Aitken, 45-64. Cambridge Mass.: Harvard University Press.
Semrau, Thorsten, and Stefan Sigmund. 2012. Networking ability and the financial performance of new ventures: A Mediation Analysis among Younger and More Mature Firms. *Strategic Entrepreneurship Journal* 6 (4): 335-54.
Shane, Scott. 2000. Prior knowledge and the discovery of entrepreneurial opportunities. *Organization Science* 11 (4): 448-69.
Simmel, Georg. 1992. Untersuchungen über die Formen der Vergesellschaftung. In *Gesamtausgabe*, vol. 11, ed. Otthein Rammstedt. Frankfurt am Main: Suhrkamp.
Stamm, Isabell, Allan Discua Cruz, and Ludovic Cailluet. 2019. Entrepreneurial Groups: Definition Forms, and Historic Change. *Historical Social Research* 44 (4): 7-41. doi: 10.12759/hsr.44.2019.4.7-41.
Steyaert, Chris. 2007. 'Entrepreneuring' as a conceptual attractor? A review of process theories in 20 years of entrepreneurship studies. *Entrepreneurship & Regional Development* 19 (6): 453-77.
Tajfel, Henri. 1982. Social psychology of intergroup relations. *Annual Review of Psychology* 33 (1): 1-39.
Talaulicar, Till, Jens Grundei, and Axel v. Werder. 2005. Strategic decision making in start-ups: The effect of top management team organization and processes on speed and comprehensiveness. *Journal of Business Venturing* 20 (4): 519-41.
Taylor, Michael. 1999. The small firm as a temporary coalition. *Entrepreneurship & Regional Development* 11 (1): 1-19.

Historical Social Research
Historische Sozialforschung

All articles published in HSR Special Issue 44 (2019) 4:
Entrepreneurial Groups & Entrepreneurial Families.